Probabilistic Constraint Satisfaction: Application to Radiosurgery

Russ B. Altman, MD, PhD and Rhea Tombropoulos, Section on Medical Informatics
Stanford University Medical Center, MSOB X-215, Stanford, CA 94305-5479
altman@camis.stanford.edu, rzt@camis.stanford.edu

ABSTRACT

Although quite successful in a variety of settings, standard optimization approaches can have drawbacks within medical applications. For example, they often provide a single solution which is difficult to explain, or which can not be incrementally modified using secondary "soft" constraints that are difficult to encode within the optimization. In order to address these issues, we have developed a probabilistic optimization technique that allows the user to enter prior probability distributions (Gaussian) for the parameters to be optimized as well as for the constraints on the parameters. Our technique combines the prior distributions with the constraints using Bayes' rule. The algorithm produces not only a set of parameter values, but variances on these values and covariances showing the correlations between parameters. We have applied this method to the problem of planning a radiosurgical ablation of brain tumors. The radiation plan should maximize dose to tumor, minimize dose to surrounding areas, and provide an even distribution of dosage across the tumor. It also should be explainable to and modifiable by the expert physicians based on external considerations. We have compared the results of our method with the standard linear programming approach.

INTRODUCTION

There are currently well-developed techniques for parameter estimation and optimization that are generally applicable over a wide range of science and engineering problems [5]. Unfortunately, the application of standard optimization methods within medical applications can be problematic. First, these methods do not typically provide any sort of confidence bounds in the parameter values chosen. Second, they give no insight into the possible interdependence of parameter values. Third, they often require an entirely new calculation if new information is provided dynamically. We approach optimization within a paradigm of probabilistic constraint satisfaction that solves these problems.

Constraint Satisfaction and Optimization

Constraint satisfaction is a problem solving paradigm in which legal values for a set of variables are sought, subject to constraints on both the possible values for each variable as well as the relationship between values of different variables [7]. A constraint satisfaction problem is formulated by defining 1) a set of variables whose values are sought, 2) the set of possible values (the domain) for each variable, 3) the relationships between variables (the constraints). When the parameters are continuous variables, then the problem of finding values that satisfy the constraints optimally is isomorphic to the problem of optimizing the variables so that some error function is minimized. We have described an algorithm for probabilistic constraint satisfaction that should, therefore, be useful for general optimization [1,2]. We represent the set of possible values for a variable as continuous parametric distributions over the range of possible labels. Initially, a variable distribution is described based on some prior knowledge of the range of possible labels (assumed to be Gaussian). The constraints between variables are probability distributions over functions that depend on the structural variables (described below). As constraints are introduced, they cause changes in the probability distribution for all the variables. The resulting probability distributions are the posterior probability distributions of the values for all variables.

Radiosurgery

Radiosurgery is a method for ablating brain tumors with high intensity radiation [3,4,6,8,9]. The main challenge in radiosurgery is to develop dosing schemes for which the dose to tumor is maximal and homogenous, while the dose to surrounding healthy tissue (especially tissues involved in critical functions or which are very sensitive) is minimal. With the advent of high performance robotic arms to position the xray source, it has become possible to plan radiosurgical procedures in which the xray beam impinges upon the tumor from multiple angles, so that healthy tissues receive low doses compared to the tumor. The problem of maximizing dose to the tumor and minimizing dose to healthy tissues becomes one of setting the strengths assigned to a set of xray beams impinging upon the tumor from different angles. The precise models for determining dose to a region [8] are too expensive to use for optimization of beam intensities, but the dose through a region, **S**, can be roughly estimated as the sum of the intensities of the beams which pass through the region (see Equation 11).

In general, the dose through tumor tissue should exceed some critical value, and the dose through normal tissue should be less than some safe value. The goal of an optimization program is to find values of the individual beam intensities that satisfy these constraints. Schweikard et al have reported a system which uses standard linear programming to solve this problem. It has been shown to perform well [8,9]. It is subject, however, to the pitfalls of standard optimization. It may

0195-4210/94/$5.00

not always converge, does not provide information about the range of intensity values compatible with the dosing goals, and provides no information about which part of the optimization is difficult. We have therefore attempted to replicate its performance using a system that maintains second order information about the parameter values, such as variance and covariance of beam intensities.

MATHEMATICAL FORMULATION

We represent an optimization problem as a vector of parameters (each parameter is a node in the constraint network). Each parameter has a prior probability distribution over the range of possible variable values. In our current implementation, we represent each distribution by its first two moments, the mean and variance. The mean values of each parameter are stored in a state vector, **x**. For a structural model with M parameters, the state vector is:

$$\mathbf{x} = [x_1 \quad x_2 \quad x_3 \quad . \quad . \quad . \quad x_M] \qquad [1]$$

The second moment of the state vector is stored in a matrix. The diagonal elements of the matrix contain the variances of each parameter. The off-diagonal elements of the matrix contain the covariances between parameters.

$$\mathbf{C(x)} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdot & \sigma_{x_1x_M} \\ \cdot & \sigma_{x_2}^2 & & \cdot \\ \cdot & & \cdot & \cdot \\ \sigma_{x_Mx_1} & \cdot & \cdot & \sigma_{x_M}^2 \end{bmatrix} \qquad [2]$$

If we have no information about the relationships between parameters, then the covariances are zero. As information about the relationship between parameters is gathered (as described below), the covariances may become non-zero.

Taken together, the mean vector, **x**, and the covariance matrix, **C(x)**, represent an uncertain estimate of the possible parameter values. In general, two parameters may have a complicated functional relationship. The covariance is simply a linearization of this relationship that specifies whether the values of the two parameters are correlated. Although a primitive summary of potentially complicated dependencies, the covariance is often sufficient (especially using iterative techniques to reduce estimation error) for capturing important parameter relationships.

Constraints involving the elements of the state vector, **x**, can be used to update the parameter estimates within **x** and **C(x)**. A constraint, **z**, is represented with two components; a deterministic function, **h(x)** that specifies how the value of the measurement depends on (and can be calculated from) the structural parameters in **x**, and a stochastic component, **v**, that specifies the uncertainty or tolerance in the information.

$$z = h(x) + v \qquad [3]$$

In this work, **v** is assumed to be distributed normally with mean zero. We have found that inequality constraints (in which the equality of Eq. 3 becomes an inequality) can often be approximated with appropriate Gaussian distributions.

Given a new constraint, the parameter estimates can be updated using Bayesian measurement update formulae, as is used in the Kalman Filter [5]. These formula introduce non-zero covariances between all parameters that are involved in the constraint. New values for **x** and **C(x)** are calculated as follows:

$$x_{new} = x_{old} + K\left[z - h(x_{old})\right] \qquad [4]$$

$$C(x_{new}) = C(x_{old}) - KHC(x_{old}) \qquad [5]$$

where

$$K = C(x_{old})H^T[HC(x_{old})H^T + C(v)]^{-1} \qquad [6]$$

and

$$H = \left.\frac{\partial h(x)}{\partial x}\right|_{x_{old}} \qquad [7]$$

Simply put, the new estimate for **x** is based on the old estimate plus a weighted difference between the observed and predicted value of **z**. The weighting factor, **K**, is proportional to the ratio of the uncertainty in the current estimate and the uncertainty in the constraint.[1] If **h(x)** is nonlinear in **x**, then errors in the linearization of **h(x)** can be reduced by using a modified update formula [5].

These update formulae use the information in the covariance matrix to make concerted changes in parameter values. Thus, introducing a constraint that provides information about parameter x_i, will modify the value of variables which are highly correlated with x_i. However, if the relationship between parameters is nonlinear, then the covariances may not lead to precisely correct new values. We have described an iterative procedure to effectively minimize these errors [1].

With the measurement update equations, we can solve a constraint satisfaction formulation of optimization.

[1]Note that the term within the reciprocal is a first order estimate of the uncertainty in the measurement as predicted by the model, since the variance of z, $C(z) = C(h(x) + v) = C(h(x)) + C(v) \approx HC(x)H^T + C(v)$.

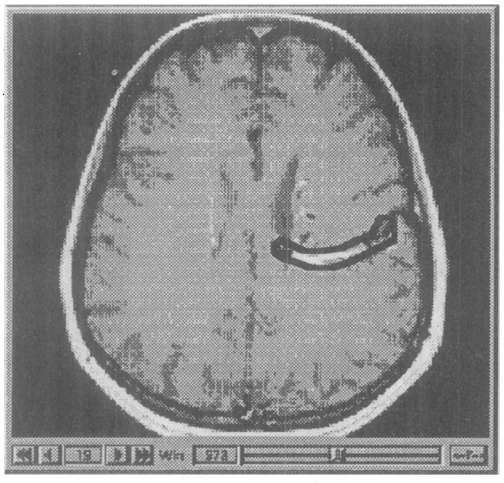

Figure 1. CT slice of tumor (middle right) used in testing.

A starting set of parameters is generated (based on previous experience or with a broad, uninformative prior distribution) and measurements are introduced (serially or in one large vector of values) to update the estimates of mean and variance. The final values of **x** and **C(x)** represent the posterior distribution of the parameter values that satisfy the constraints. We define the residual error for each constraint as follows:

$$E_i = \frac{h(x_{new}) - z_i}{\sqrt{\sigma^2_{v_i}}} \qquad [9]$$

where E_i is the error for an individual constraint, **z** is the measured value, $\mathbf{h(x_{new})}$ is the predicted value based on the best estimate, and σ is the variance of $\mathbf{v_i}$.

APPLICATIONS TO RADIOSURGERY

For radiosurgery, the vector to be estimated, **x**, contains the values of the beam weights for *N* beams that impinge on the tumor mass:

$$x = \begin{bmatrix} w_1 & w_2 & w_3 & \cdots & w_N \end{bmatrix} \qquad [10]$$

We generate one constraint for every volume element through tumor or sensitive tissue. If there are M such volume elements, then a constraint for the individual volume element through which a subset S of the N beams pass is represented as:

$$D_S = \sum_{i \in S} w_i + v \qquad [11]$$

D_S, is the target dose for the volume element (based on its status as tumor or normal). **v** is a random variable describing the tolerance we have for variation in the value of D_S. Given a starting value for the parameter vector, and all the constraints, we can calculate an improved estimate of the parameter values.

Example and Evaluation

We illustrate the performance of our method on a tumor that is shown in Figure 1. The irregularity of this tumor makes it a good test of the ability to refine the beam intensities. As a control, we applied the standard linear programming algorithm. For this tumor, there are only two types of tissues: normal and tumor. The linear programming technique finds a set of beam weights which produce a dose to all tumor tissues between 2000 and 2300 R (100 Rad = 1 Grey), while minimizing total dose to normal tissue. We translated these constraints into normal distributions; tumor tissues were constrained to have a mean dose of 2100 R with a variance of 1000 R^2.

The initial mean values for the beam weights were set to 150 R with a variance of 1600 R^2. The mean value is based on an estimate of how many beams travel through the tumor, and what the average dose to the entire tumor should be. The variance is chosen to allow a range of values, while minimizing the possibility that the program chooses a negative weight for a beam (a physical impossibility). The computational procedure was as follows:

1. Using the starting parameter estimates, all constraints were used to update the parameter estimate.
2. Any beam weights that were negative after step 1, were set to 0, and removed from the optimization.
3. The overall satisfaction of constraints is measured, using Equation 9. If there is no change in the result (within a user defined tolerance), then the algorithm halts.
4. If the residual errors are still large, the remaining beam weights were retained and used for another round of updating with the constraints (that is, loop to step 1).

Evaluation

In order to evaluate the internal consistency of our result, we analyzed the the distribution of errors for all constraints. The average error for all constraints is 1.57 SD. The error distribution assures us that the optimization was able to find a solution that is reasonable with respect to the provided constraints.

A more significant evaluation of our result entails a comparison with the "gold standard" of linear programming. There are two parameters with which we compared the methods: ability to deliver dose specifically to the tumor (that is, match the contour of the tumor precisely and minimize dosage to surrounding normal tissue), and ability to deliver tumor dose homogenously (that is, all tumor regions get approximately the same dosage). Figure 2 compares the ability of linear programming and probabilistic optimization to deliver adequate radiation to tumor cells. Figure 2AB shows for each of the two methods, the three-dimensional contour of volume elements that receive least 50% of the maximum

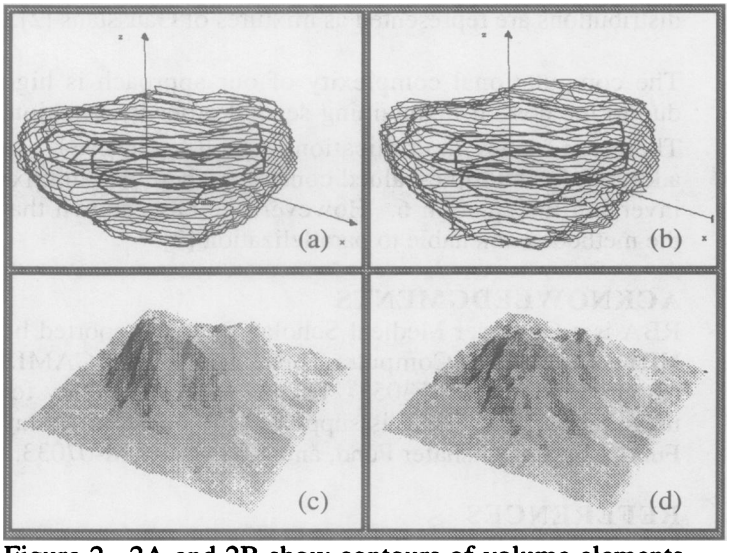

Figure 2. 2A and 2B show contours of volume elements receiving at least 50% of maximum dose. They each contain the contour of the tumor, as defined by an expert. 2C and 2D show the tumor dose on the slice of tumor shown in Fig 1. for each method. (2B/2D = our method).

---

dose delivered to any volume element. Superimposed upon these contours are the tumor contour defined manually be an expert physician. Ideally, all tumor volume elements will have nearly 100% maximum dosage, and the drop off would be precipitous. In practice, we examine the way in which the 50% contours match the manually segmented tumor boundary to get a feeling for the rate at which the tumor dose falls at the edges of the tumor. Figure 2CD demonstrates, for a single section through the tumor, how the incoming xray beams concentrate the dose within the tumor.

The homogeneity of dose distribution is illustrated by Figures 2 and 3. In the ideal, Figure 2CD would have a flat plateau in the area of tumor, and a steep drop off to zero for the surrounding area. Both linear programming and probabilistic optimization provide quite good approximations to a plateau. Figure 3 plots the volume of tumor exposed to increasingly greater percentages of the maximum dose. For example, it shows that for both results virtually the entire tumor (of volume 3800 mm^3) receives at least 80% of the maximal dose received by any tumor volume element, but that this ratio falls off rapidly so that only about half the volume elements (2000 mm^3) receive 90% or more of the maximal dose. Only about 15% of voxels receive 95% or more of the maximal dose. The integrated area of the curves in Figure 3 is a measure of the submaximal radiation or *efficiency*. For a perfect dose scheme, the integral should be 100. Both the linear programming and constraint satisfaction approaches produce efficiencies of 88%.

We conclude that the probabilistic optimization provides solutions that are compatible with linear programming

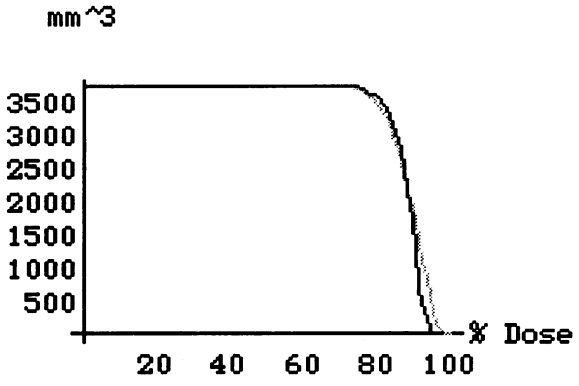

Figure 3. The number of tumor volume elements (mm^3) receiving a percent of the maximum dose (black is our method). The tumor volume is 3800 mm^3 and virtually all the tumor receives at least 80% of the maximum dose.

---

methods. We have now run other comparisons (not reported here) that produce similar agreement. The chief advantage of our method is the additional information provided by the probabilistic optimization. First, we have explicit confidence in the values for each beam strength. Figure 4 shows a plot of distribution of variances for all beams. These variances provide specific information that can be used to understand which beam values are critical to the dosage. For example, most of the beams have a variance between 10 and 40, indicating that these beams can be adjusted 3 to 7 units without significant effect on the final dose. Some beams, however, have variances up to 80 or more, which indicates significantly more flexibility for adjustment.

Secondly, the covariance matrix provides information about the correlated beam strengths. This allows us to identify key subsets of interacting beams (perhaps across multiple constraints, and not apparent by a scan of individual constraints). For example, Figure 5 plots the largest covariances within the final set of nonzero beams. This information suggests which clusters of beams (and the tumor regions they affect) are tightly linked.

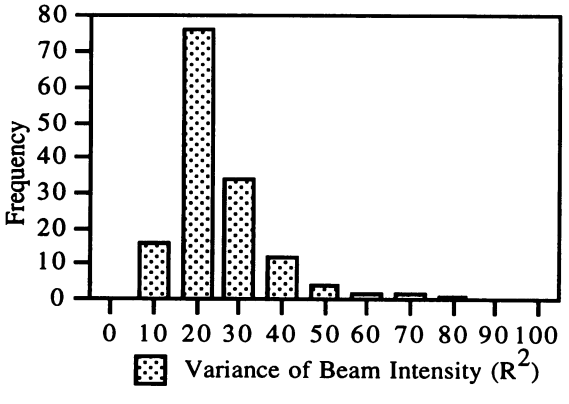

Figure 4. Histogram of the variance of beam intensities for 152 final nonzero beams. Most beams are defined tightly (with variance of 20 R^2), but some have more freedom to change values.

DISCUSSION AND CONCLUSIONS

These results demonstrate that a probabilistic constraint satisfaction formulation can produce answers comparable with linear programming. Although we have not demonstrated any practical use of the variance/covariance information, we believe that there are clear advantages of having the second order information in the context of semiautomated systems which will allow user interaction and modification. First, new constraints can be immediately introduced using the update equations (4-7) given above. The covariance information allows all beams to be updated in a concerted fashion. Second, information about variance of beam intensities can be used to immediately recognize which beams have a narrow therapeutic range, and which can be varied within a broader set of values.

This approach to parameter estimation has proven quite versatile in a number of domains, including the estimation of macromolecular structure, model driven CT image segmentation, and automated interpretation of MRI images of the cervical spine [1]. It is possible to solve a large variety of general optimization problems within this framework: the variables make up the state vector, and the constraints are the conditions that restrain the optimization function.

The use of two moments of a distribution can be limiting: there are clearly constraints on structure that may be multimodal or in some other way non-Gaussian. Also, we may want to represent non-normal posterior distributions for our variables. The introduction of higher moments adds computational complexity. However, because of their importance for some problems, we are currently investigating ways to relax the assumption of normality in the context of massively parallel computers. We have recently reported an extension in which arbitrary distributions are represented as mixtures of Gaussians [2].

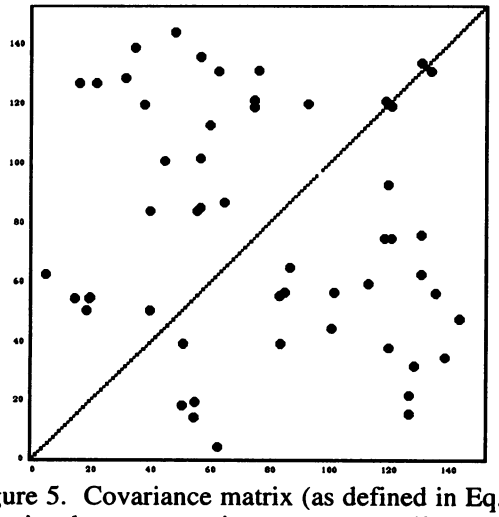

Figure 5. Covariance matrix (as defined in Eq. 2) showing largest covariances among 152 nonzero beams in the probabilistic solution.

The computational complexity of our approach is high due to the cost of maintaining second order information. There are matrix multiplications which are $O(N^3)$. In addition, for multiple-valued constraits, there is a matrix-inversion in Equation 6. However, we have shown that the method is amenable to parallelization [2].

ACKNOWLEDGMENTS

RBA is a Culpeper Medical Scholar, and is supported by NIH LM-05652. Computers provided by the CAMIS resource, NIH LM-05305. Thanks to J.F. Brinkley for useful comments. RZT is supported by the Sheik Enany Fund, Lorraine Ulshafer Fund, and NIH grant LM-07033.

REFERENCES

[1] Altman, R. B. *A probabilistic algorithm for calculating structure: Borrowing from simulated annealing.* Proceedings, Ninth Conference on Uncertainty in Artificial Intelligence, Washington, D.C., July 1993, Morgan Kaufmann Pub., San Mateo, CA pp. 23-31.

[2] Altman, R.B., Chen, C.C., Poland, W.B., Singh, J.P. *Probabilistic constraint satisfaction with non-gaussian constraint noise.* Proceedings, Tenth Conference on Uncertainty in Artificial Intelligence, Seattle, WA, August, 1994, Morgan Kaufmann Pub., San Mateo, CA, pp. 15-22.

[3] Betty, O.O., Munari, C., and Rosler, R. Stereotactic Radiosurgery with the Linear Accelerator: Treatment of Arteriovenous Malformations. Neurosurgery, 24(3): 311-321, 1989.

[4] Brahme, A. Optimization of stationary and moving beam radiation therapy techniques. Radiother. Oncol., 12:127-140, 1988.

[5] Gelb, A., *Applied Optimal Estimation.* 1984, Cambridge, Massachusetts: MIT Press.

[6] Larsson, B., et al. The High Energy Proton Beam as a Neurosurgical Tool. Nature, 182:1222-3, 1958.

[7] Mackworth, A.K. *Consistency in networks of relations.* Artificial Intelligence. 8, 99-118 (1977).

[8] Schweikard, A., Adler, J.R., and Latombe, J.C., Motion Planning in Stereotaxic Radiosurgery. Proc. IEEE Int. Conf. Robotics and Automation, Atlanta, CA, May 1993, 1909-1916. (Extended version to appear in IEEE Tr. Robotics and Automation.)

[9] Schweikard, A., Tombropoulos, R.Z., Adler, J.R., and Latombe, J.C., Treatment planning for a radiosurgical system with general kinematics. in press, Proc. IEEE Int. Conf. Robotics and Automation, 1994.